HAL
open science

Implementation strategies of a contract-based MRI examination reservation process for stroke patients

Na Geng, Xiaolan Xie, Zhibin Jiang

To cite this version:

Na Geng, Xiaolan Xie, Zhibin Jiang. Implementation strategies of a contract-based MRI examination reservation process for stroke patients. European Journal of Operational Research, 2013, 231 (2), pp.371-380. 10.1016/j.ejor.2013.05.043 . hal-00911127

HAL Id: hal-00911127

https://hal.science/hal-00911127v1

Submitted on 2 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Implementation Strategies of A Contract-based MRI Examination Reservation Process for Stroke Patients

Na GENG[a], Xiaolan XIE[a,b], Zhibin JIANG[a]

[a] Department of Industrial Engineering & Logistics Management
Shanghai Jiao Tong University
800 Dongchuan Road 200240 Shanghai, China

[b] Center for Biomedical and Healthcare Engineering & LIMOS-ROGI CNRS UMR 6158
Ecole Nationale Supérieure des Mines de Saint Etienne
158 Cours Fauriel, 42023 Saint Etienne, France

Submitted on 07/02/2012. Revised: 03/15/2013, 05/22/2013

Corresponding author:
Prof. Xiaolan XIE

Abstract: Timely imaging examinations are critical for stroke patients due to the potential life threat. We have proposed a contract-based Magnetic Resonance Imaging (MRI) reservation process [1] in order to reduce their waiting time for MRI examinations. Contracted time slots (CTS) are especially reserved for Neural Vascular Department (NVD) treating stroke patients. Patients either wait in a CTS queue for such time slots or are directed to Regular Time Slot (RTS) reservation. This strategy creates "unlucky" patients having to wait for lengthy RTS reservation. This paper proposes and analyzes other contract implementation strategies called RTS reservation strategies. These strategies reserve RTS for NVD but do not direct patients to regular reservations. Patients all wait in the same queue and are served by either CTS or RTS on a FIFO (First In First Out) basis. We prove that RTS reservation strategies are able to reduce the unused time slots and patient waiting time. Extensive numerical results are presented to show the benefits of RTS reservation and to compare various RTS reservation strategies.

Keywords: healthcare, hospital, capacity allocation, capacity reservation, implementation strategies

I. INTRODUCTION

This paper is motivated by our collaboration with a large French university teaching hospital in order to reduce the waiting time of stroke patients treated in Neural Vascular Department (NVD). A stroke is the sudden loss of brain function, which is usually caused by interruption of the blood supply (ischemic stroke) and rupture of a blood vessel (hemorrhagic stroke). The brain is deprived of oxygen and nerve cell death will occur in the area of the brain with no blood flow. The patients will suffer from severe physical and cognitive deficits. It is crucial for stroke patients to have timely imaging examinations in order to have appropriate diagnosis and treatment.

However, significant delays are observed because many key examinations rely on expensive and heavily used imaging facilities such as Magnetic Resonance Imaging (MRI). In France, the average waiting time for MRI examinations is about 30~40 days. Stroke patients, just like the other routine patients, have to reserve the time slots via fax or via telephone for the emergency patients. Long waiting time has a negative impact on quality of care and patient service [2].

In order to reduce the waiting time of stroke patients, we proposed in [1] a contract-based MRI examination reservation process. The imaging department managing the MRI facilities reserve each week some Contracted Time Slots (CTS) for NVD treating stroke patients. Stroke patients can be served by either CTS or regular time slots (RTS) in case of arrival surges of stroke patients. CTS decisions and RTS assignment policy determine the efficiency of the reservation process. The former is the number of CTS and its distribution over the time, whereas the latter refers to the policy for assigning incoming patients to RTS. We proposed a method combining stochastic programming model and Markov Decision Process (MDP) to simultaneously determine the two decisions. Structural properties of the optimal RTS assignment policy were established by an average cost MDP model. It is proved that there exists a threshold L_i for each day i and the optimal RTS assignment control consists in sending patients to RTS by keeping CTS queue length below L_i. Then Monte Carlo approximation and local improvement were used to determine CTS decisions and RTS assignment policy. Computational results showed that the reservation process can greatly reduce the average waiting time of stroke patients. However, there exist some “unlucky” patients who are directed to RTS and have to wait much longer time than those who wait for CTS.

In order to solve this problem, this paper proposes three new MRI reservation processes. In

these processes, RTS is reserved for NVD and added to the list of CTS. CTS and RTS time slots are grouped according to their day of availability and filled by patients. No patient is directed to wait for lengthy RTS. All patients wait in the same queue for MRI time slots and are scheduled to either CTS or RTS in a FIFO (First In First Out) order and according to the release dates of CTS and RTS.

Three RTS reservation policies are proposed in this paper based on patient queue length information and some service ratio criterion. We first prove that RTS reservation policies outperform RTS assignment policy with shorter patient waiting time and less unused MRI time slots. We then perform extensive numerical experiments to compare the performances of different contract implementation strategies.

The rest of this paper is organized as follows. Literature review is described in Section II. Section III defines and compares different MRI-contract implementation strategies. Formal proofs that RTS reservation strategies improve RTS assignment strategy are given in the Appendix. Section IV presents computational results to show the efficiency of RTS reservation strategies and the impact of different problem parameters. Conclusions and perspectives are given in Section V.

II. Literature Review

Capacity reservation and scheduling of imaging facilities, such as computer tomography (CT) and MRI scanners, have received limited coverage. The two earliest contributions are [3] and [4]. Simulation was used in [3] to investigate the effect of scheduling rules on patient waiting time and physician idle time for X-rays in a chest radiology department. It was demonstrated in [4] that more technicians and orderlies do not imply better service quality. To improve the radiology services, the emphasis was the design of the management systems and scheduling techniques.

A new scheduling method was proposed in [2] to identify improvement potentials in order to reduce MRI access time. Finite-horizon dynamic program was used in [5] to effectively allocate the expensive imaging diagnosis capacity among several classes of patients during a day. Properties of the optimal policies are identified in order to design the outpatient appointment schedule, and establish dynamic priority rules for admitting patients into services. [6] proposed a simple approach for dividing the available diagnosis capacity between emergency and inpatients on the one hand and outpatients on the other. MDP was used in [7] for the admission of multi-priority patients on a waiting queue to a diagnostic resource. An

approximate dynamic programming approach was proposed to overcome the state space explosion problem. The same problem was solved in [8] with protection level policies by protecting a part of the capacity from the lower priority jobs in order to make it available for the future higher priority jobs. A perturbation analysis technique was proposed to evaluate sample-path gradients with respect to the protection levels. A stochastic approximation approach was used to determine the optimal protection levels.

MDP was used in [9] to allocate two CT-scanners to three patient groups with different arrival patterns and different cost-structures in order to maximize the total expected reward. [10] determined the optimal number of outpatients to schedule and the assignment of outpatients to a variable-block/fixed-interval appointment schedule. An MDP approach was proposed in [11] to decide whether to accept requests for MRI examinations from patients with different priorities such as inpatients and outpatients. Different examination types, cancellations, no-shows and over-booking, and same-day demand were considered. A continuous-time Markov decision process was used in [12] to model the problem of accepting or rejecting the reservation of different services by different classes of customers. The solution strategy was proposed by using simulation-based approximate dynamic programming (ADP) combined with a discrete event simulation of the service period. Numerical experiments show that the heuristic ADP algorithm performs very well in terms of objective function value, solution time, and memory requirements.

Queueing theory was used in [13] to determine the number of schedule slots to open in a radiology department for urgent CT and ultrasonography in order to keep the rescheduling rate of routine studies to accommodate emergencies below a certain level. [14] examined a multi-period capacity allocation model with upgrading by considering multiple product types and multiple classes of demand. The optimal allocation policy was shown to be a simple two-step algorithm: meet demands with available capacity of the same-class, and then upgrade customers until a protection limit of a class-dependent capacity. Bounds of optimal protection limits were proposed.

Most existing studies focus on medical service capacity allocation and scheduling from the perspective of service provider side except for [1, 15, 16]. [1] proposed contracted based MRI reservation process, and determined CTS decisions and RTS assignment policy by combining stochastic programming model and average-cost MDP. It differs from the previous studies on critical resource scheduling by investigating the problem from a totally different perspective and explores solutions from the perspective of a given class of customers, i.e., stroke patients.

CTS are pre-reserved for stroke patients by taking their average waiting time and unused CTS into account. On this basis, [15, 16] considered the possibility of canceling some CTS and proposed average-cost MDP approaches. [15] determined simultaneously RTS assignment policy and one-day advance cancellation policy for CTS; whereas [16] determined the RTS assignment policy, one-day and two-day advance cancellation policy for CTS. This paper investigates different strategies for implementing an MRI-contract in order to reduce waiting time variation. Advance cancellation of CTS is not considered in this paper.

III. Problem and Strategies

A. MRI time slot contracting and reservation

In order to reduce the waiting times of stroke patients, we proposed in [1] a contract-based MRI examination reservation process. In this reservation process, NVD treating stroke patients is allocated a certain number of time slots every day which is called Contracted Time Slots (CTS). The distribution of CTS over time will be called the **contract.** When the queue for CTS is too long, stroke patients can also use additional time slots via regular reservation which is called Regular Time Slots (RTS). Patients either wait for MRI time slots in a CTS queue or are directed to RTS via lengthy regular reservation process (30-40 days). This strategy will be called **RTS assignment policy**. Although the combination of contract and RTS assignment policy greatly reduces the average waiting time of stroke patients, the variation of patients' waiting time is quite large. The "unlucky" patients assigned to RTS have to wait much longer than those waiting for CTS.

In order to improve waiting time distribution, this paper proposes three new implementation strategies. These new strategies still make use of the same contract. With these new strategies, no patient is directed to wait for RTS reservation, and all patients wait in the same queue for MRI time slots. Additional MRI time slots are reserved by NVD through the regular reservation process. The reservation depends on the queue length of patients, and the corresponding time slots will be available in T time periods later. These additional time slots and CTS are pooled together and given to strokes patients on a FIFO basis. These new reservation processes are called **RTS reservation policies**.

B. System dynamics under contract implementation strategies

In this paper, the following assumptions are made:

A1. Only MRI examination is considered and each patient can use either one CTS or one

RTS.

A2. Only one class of patients is considered and all patients have equal priority.

A3. The arrival of patients varies during a week but is stationary from one week to another. Further, the number of arrivals in one day is independent of the arrivals of other days.

A4. The contract is known and remains the same from one week to another.

A5. The waiting time for a time slot via regular reservation is a constant T.

Remark 1: The average waiting time of the stroke patients served by RTS is approximated by a constant T in this paper. In practice, the use of contract-based solution is expected to modify T and reserving more CTS for stroke patients will lead to longer T. However, assumption A5 is reasonable for the following reasons. First, in practice MRI time slots are shared among multiple departments. The proportion of MRI usage by NVD is usually not too high. For example, the hospital we collaborated with has 5 MRI machines shared by 61 medical services of the hospital, many external services and the imaging department itself for various diagnosis, researches and teaching activities. The change of reservation process of NVD alone is expected to have limited impact on the delay of regular reservation. Second, our simulation studies show that change of waiting time for regular reservation due to the use of well-designed contract is quite limited when the overall workload of MRI facilities is reasonable and is for example below 90%. Third, the impact of CTS on T can be further reduced by appropriate advance CTS cancellation strategies not addressed in this paper but investigated in [15] and [16]. □

The MRI reservation problem is characterized by the following notation:

t index of days with $t=1,2, \ldots;$

T number of days for obtaining a time slot through regular reservation with $T > 1$;

a_t number of patients arrived in day t. By assumption 3, daily arrivals a_t are mutual independent random variables and weekly arrivals (a_{7w+1}, a_{7w+2}, …, a_{7w+7}) are identically distributed for all $w = 0, 1, \ldots$. As a result, the arrival process is characterized by probability matrix $[r_{ij}]$ for $i = 1, \ldots, 7$ and for all $j \geq 0$ with r_{ij} denoting the probability of j arrivals in day i;

n_t number of CTS of day t. By assumption 4), $n_t = n_{t+7}$ and **the contract** can be denoted by a 7-integer vector $\boldsymbol{n} = [n_1, \ldots, n_7]$;

c penalty factor of an unused MRI time slot. It serves as a weighting factor in order to balance waiting time and unused MRI time slots.

The cost structure is as follows. Each patient incurs a cost of 1 for each waiting day for either CTS or RTS. Each unused CTS or RTS time slot incurs a cost of c. The goal is to minimize the long run average cost incurred by patient waiting and unused time slots of both CTS and reserved RTS.

Four implementation strategies $\mathbf{P_i}$ with i =0, 1, 2, 3 will be considered in this paper. Each policy $\mathbf{P_i}$ is associated with the following notation:

$y_{i,t}$ number of patients directed to RTS or number of RTS reserved at the end of day t;

$u_{i,t}$ number of unused time slots in day t;

$x_{i,t}$ total number of patients waiting for a time slot at the end of day t including those directed to RTS but not yet served. It will be called **global queue length**;

$d_{i,t}$ number of patients having received their time slots and hence left in day t.

The capital letter of each notation denotes the cumulative total from 0 to t. Notation A_t, $D_{i,t}$ and $U_{i,t}$ will be used. The following notation is also used:

$$q_{i,t} = x_{i,t} - Y_i\left(t-T+1, t-1\right) - y_{i,t} \tag{1}$$

where $Y_i(t',t) = y_{i,t'} + \ldots + y_{i,t}$, $Y_i(t\text{-}T\text{+}1,t\text{-}1)$ denotes the outstanding RTS assignment or reservation in day t. If $\mathbf{P_i}$ is a RTS assignment policy, $q_{i,t}$ corresponds to the **CTS queue length** and $x_{i,t} = q_{i,t} + \sum_{\tau=t-T+1}^{t} y_{i,\tau}$. If $\mathbf{P_i}$ is a RTS reservation policy, $x_{i,t}$ is the queue length of waiting patients and $q_{i,t}$ equals with $x_{i,t}$ minus the total number of outstanding RTS reservations.

With a given contract $\boldsymbol{n}$ and a given implementation strategy $\mathbf{P_i}$, the sequence of events in each day is as follows. At the beginning of the day, the queue lengths $q_{i,t\text{-}1}$ and $x_{i,t\text{-}1}$ as well as outstanding RTS assignments / reservations $y_{i,t'}$ for $t' = t - T$ *to* t-1 are known, a_t new patients arrive, n_t CTS are available. If RTS assignment is used, MIN{n_t, $q_{i,t\text{-}1}$+a_t} patients in the CTS queue are served by CTS of day t and $y_{i,t\text{-}T}$ patients are served by RTS of day t. If RTS reservation is used, $y_{i,t\text{-}T}$ additional time slots arrive and MIN{n_t+$y_{i,t\text{-}T}$, $x_{i,t\text{-}1}$+a_t} patients are served by CTS and RTS available in day t. At this moment, the decision variable $y_{i,t}$ is selected according to policy $\mathbf{P_i}$. $y_{i,t}$ patients are directed RTS and depart from the CTS queue if

RTS assignment is used and $y_{i,t}$ RTS time slots are reserved if RTS reservation is used. All queue lengths and performance indicators are then updated. Fig. 1 illustrates the dynamics of the two types of strategies.

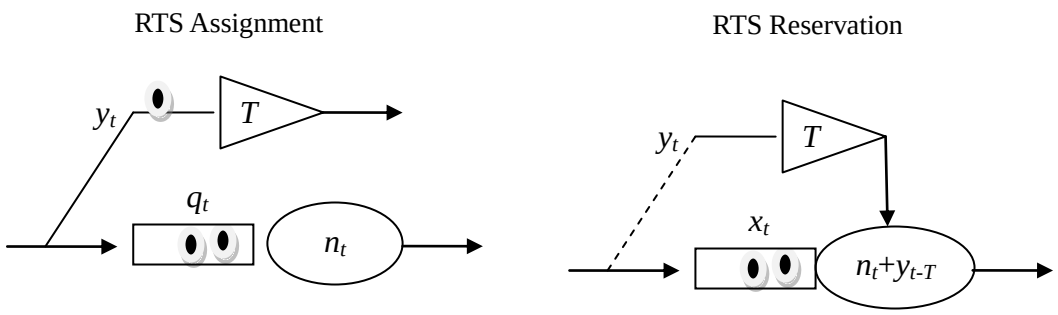

Fig. 1. RTS assignment and RTS reservation

Without loss of generality, we assume that the system starts empty with

$$\left(x_{i,t}, q_{i,t}, y_{i,t}\right) = \mathbf{0}, \forall t \le 0 \tag{2}$$

The system dynamics can be characterized by the following:

$$q_{i,t} = \left(q_{i,t-1} + a_t - n_t - y_{i,t}\right)^+, \text{ for RTS assignment} \tag{3}$$

$$x_{i,t} = \left(x_{i,t-1} + a_t - n_t - y_{i,t-T}\right)^+, \text{ for RTS reservation} \tag{4}$$

$$u_{i,t} = \left(n_t - q_{i,t-1} - a_t\right)^+, \text{ for RTS assignment} \tag{5}$$

$$u_{i,t} = \left(n_t + y_{i,t-T} - x_{i,t-1} - a_t\right)^+, \text{ for RTS reservation} \tag{6}$$

$$D_{i,t} = A_t - x_{i,t} \tag{7}$$

The performance measures considered in this paper include: long-run average cost $\bar{C}_i$, average unused time slots $\bar{U}_i$, average patient waiting time $\bar{W}_i$ and average queue length $\bar{X}_i$.

$$\bar{C}_i = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \left(c u_{i,t} + x_{i,t}\right) \tag{8}$$

The evaluation of other performance measures is straightforward. Further, by Little's law,

$$\bar{W}_i = \bar{X}_i / \bar{A} \tag{9}$$

where $\bar{A} = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} a_t = \frac{1}{7} \sum_{t=1}^{7} E(a_t)$ is the average daily arrival rate. This implies that higher global queue length $x_{i,t}$ implies higher average waiting time.

C. Contract implementation strategies

One RTS assignment policy denoted $\mathbf{P_0}$ and three RTS reservation policies $\mathbf{P_1}$, $\mathbf{P_2}$ and $\mathbf{P_3}$ are considered in this paper.

$\mathbf{P_0}$ is the optimal **RTS assignment policy** which is shown in [1] to be a policy characterized by a threshold L_t associated with each day with $L_t = L_{t+7}$. This policy keeps the CTS queue length q_{0t} at the end of each day t below or equal to L_t. As a result,

$$y_{0,t} = \left(q_{0,t-1} + a_t - n_t - L_t\right)^+ \tag{10}$$

$$q_{0,t} = \min\left\{\left(q_{0,t-1} + a_t - n_t\right)^+, L_t\right\} \tag{11}$$

With a RTS reservation policy $\mathbf{P_i}$ (i =1, 2, 3), no one is directed to regular reservation and all patients wait in the same patient queue and are served in FIFO order. At the beginning of period t, the length of the patient queue is $x_{i,t-1}$, the total number of available time slots is $n_t+y_{i,t-T}$. The number of RTS to reserve $y_{i,t}$ is determined via different methods.

$\mathbf{P_1}$ is a RTS reservation policy such that $y_{1,t} = y_{0,t}$. More precisely, $\mathbf{P_1}$ keeps track of an artificial CTS queue length $q_{0,t}$ as if $\mathbf{P_0}$ were used and it determines $y_{0,t}$ and hence $y_{1,t}$ with the artificial queue and relation (10). That is, $\mathbf{P_1}$ is similar to $\mathbf{P_0}$ but with additional RTS reserved instead of sending patients directly to RTS. In the following, $\mathbf{P_1}$ is called **RTS reservation with artificial queue**.

$\mathbf{P_2}$ is a RTS reservation policy defined as follows:

$$y_{2,t} = \left(x_{2,t} - Y_2\left(t-T+1, t-1\right) - L_t\right)^+ \tag{12}$$

and is called **RTS reservation with real queue**. The number $y_{2,t}$ of RTS to reserve is determined by considering its effect on patient queue. From assumption A5, the RTS reservation decision y_t only impacts on the patient queue in day $t+T$. $\mathbf{P_2}$ tries to keep the expected queue length at time $t+T$ as close as possible to a threshold, i.e. $E\left(x_{t+T}\right) \approx L_t^{'}$ which

implies the following relation

$$E\left(x_t + \sum_{s=t+1}^{t+T}\left(a_s - n_s - y_{s-T} + u_s\right)\right) \approx L'_{t+T}$$

This relation is further approximated by neglecting unused time slots

$$x_t - Y\left(t-T+1, t-1\right) - y_t \approx L'_{t+T} + \sum_{s=t+1}^{t+T}\left(n_s - E\left(a_s\right)\right)$$

With $L_t = L'_{t+T} + \sum_{s=t+1}^{t+T}\left(n_s - E\left(a_s\right)\right)$ and the nonnegativity of y_t, the above relation leads to (12). Combining with (12) with (1),

$$q_{2,t} \le L_t \tag{13}$$

P$_3$ is also a RTS reservation policy but is service ratio-oriented and will be called **RTS reservation with service ratio** α. Here the service ratio at the end of a day is defined as the probability that all existing patients are served. Each day *t*, the number of RTS to reserve is determined such that the service ratio at the end of day *t+T* is at least α, i.e.

$$P\left(x_{3,t+T-1} + a_{t+T} \le n_{t+T} + y_{3t}\right) \ge \alpha$$

This policy requires the determination of the probability distribution of $x_{3,t+T-1} + a_{t+T}$ which depends on the current queue length and all outstanding RTS reservations. In this paper, this probability is determined by Monte Carlo simulation from day *t+1* to *t+T* with a fixed but large number of random samples of patient arrivals.

Remark 2: From the description of the problem, the dynamics of an RTS reservation policy can be modeled as a Markov chain but its state space is huge and beyond practical limit as the state at the beginning of each day includes not only the length of the patient queue but also the number of RTS reserved in each of the *T*-1 previous days, i.e. each state corresponds to a *T*-dimension vector. Direct analysis of the underlying Markov chain is impossible. For this reason, we compare the implementation strategies by sample path comparisons and simulation studies. □

D. Performance comparison of contract implementation strategies

This sub-section compares policies **P$_0$**, **P$_1$** and **P$_2$** under the condition of common arrivals and with the same thresholds L_t for all policies. Sample-path comparison is used in the Appendix

to establish the results in Table 1.

From Table 1, $\mathbf{P_1}$ and $\mathbf{P_2}$ both improve the RTS assignment policy $\mathbf{P_0}$. More specifically, the global queue length under $\mathbf{P_0}$ is the longest one, whereas that under $\mathbf{P_1}$ is the shortest, among policies $\mathbf{P_0}$, $\mathbf{P_1}$ and $\mathbf{P_2}$. This together with (9) implies $\bar{W}_0 \geq \bar{W}_2 \geq \bar{W}_1$. In terms of additional time slots reserved, $\mathbf{P_2}$ uses the least number of additional time slots, whereas $\mathbf{P_0}$ and $\mathbf{P_1}$ use the same number of additional time slots for every period *t*. Therefore, $\mathbf{P_2}$ has the least number of unused time slots while $\mathbf{P_1}$ and $\mathbf{P_0}$ have the same average unused time slots. With the longest waiting time and highest number of unused time slots, $\mathbf{P_0}$ has the highest cost.

Table 1: Performance comparison of policies $\mathbf{P_0}$, $\mathbf{P_1}$ and $\mathbf{P_2}$

Performance indicator	**Policy ordering**
Global queue length	$x_{0,t} \geqslant x_{2,t} \geqslant x_{1,t}$
Reservation of additional time slots	$y_{2,t} \leqslant y_{1,t} = y_{0,t}$
Ratio of unused time slots	$\bar{U}_2 \leq \bar{U}_1 = \bar{U}_0$
Average waiting time	$\bar{W}_0 \geq \bar{W}_2 \geq \bar{W}_1$
Average cost	$\bar{C}_0 \geq \bar{C}_1,\ \bar{C}_0 \geq \bar{C}_2$

IV. Experimental Study

This section presents numerical experiments performed to evaluate the performance of several MRI reservation processes. We then perform sensitivity analysis to show how the reservation processes depend on different factors such as *T* and *c*, and patient arrival rates.

A. Numerical experiments

We first describe the base case corresponding to our real case study. From the data collected from the NVD of our study, the average numbers of patient arrivals during the week are as follows: {1, 0.89, 0.95, 1.16, 1.53, 0.16, 0.05} which sum up to 5.74 patients per week. The number of patient's arrival each day is assumed to follow a Poisson distribution. The average waiting time for RTS is in the range of 30~40 days with an average of T = 35 days. The weight, *c*, is set to 15. The base case is then modified to investigate the impact of parameters *c*, *T*, and patient arrival rates.

The control parameters of the four policies are determined as follows. For any contract **n**, the optimal thresholds of $\mathbf{P_0}$ can be determined by linear programming as shown in [1] and are denoted $\mathbf{L_0}$. For $\mathbf{P_3}$, a line search with golden point insertion in interval [0, 1] is used to determine the optimal service ratio α. Two parameter settings are considered for $\mathbf{P_1}$ and $\mathbf{P_2}$.

$\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ are policies $\mathbf{P_1}$ and $\mathbf{P_2}$ with thresholds $\mathbf{L_0}$. Another threshold vector is also derived with local search (increasing or decreasing one CTS in one day) to improve $\mathbf{P_1}$ and $\mathbf{P_2}$ by starting from $\mathbf{L_0}$. Let $\mathbf{L_1}$ and $\mathbf{L_2}$ be the corresponding threshold vectors, and $\mathbf{P_1L_1}$ and $\mathbf{P_2L_2}$ be the corresponding policies.

A common randomly generated sample path **a** of patient arrivals over a time horizon of 10000 days is used in control parameter optimization of policies $\mathbf{P_3}$, $\mathbf{P_1L_1}$ and $\mathbf{P_2L_2}$. All policies are then evaluated by ten replications of the simulation of the MRI reservation process over 10000 days. Further, with policy 3, at the end of each day, 1000 replications of the on-line simulation over the next T days are performed in order to estimate the patient queue length distribution at the end of day $t+T$.

B. RTS assignment vs. RTS reservation

This section considers the base case and compares the RTS assignment policy $\mathbf{P_0}$ and RTS reservation policies P_1L_0 and P_2L_0 in order to confirm theoretical results of Section IV. The contract under consideration is $\boldsymbol{n_0}$ = [1,1,1,1,2,0,0] which is the optimal policy with 6 CTS. Remind that the average weekly demand is 5.74. The optimal thresholds of P_0 are L_0 = [6,6,6,6,5,6,6].

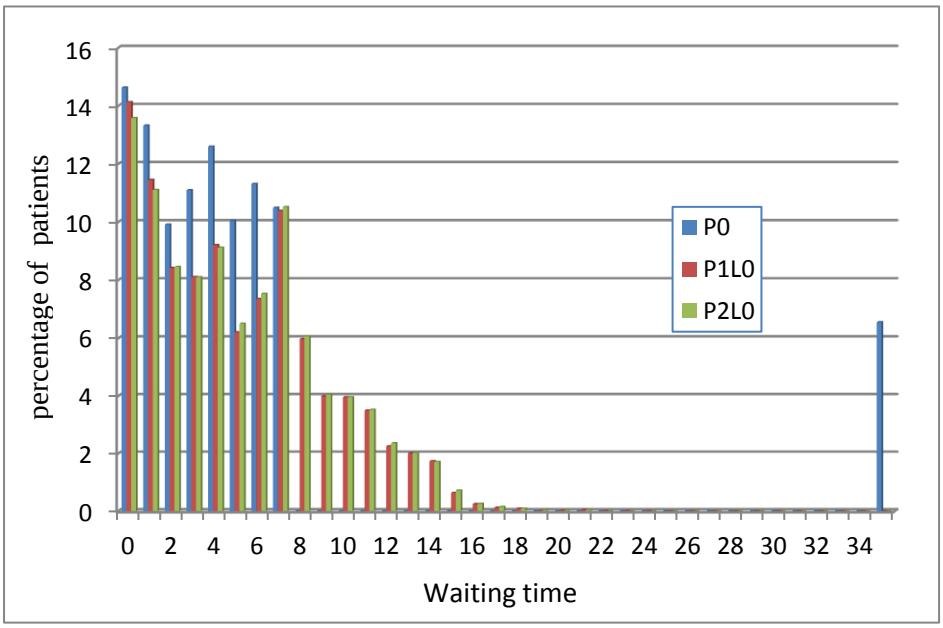

Fig. 2. Waiting time distribution

Fig. 2 gives the waiting time distributions of the three policies. The $\mathbf{P_0}$ leads to waiting times of either 0 to 7 days or 35 days with 5.56% of unlucky patients having to wait 35 days. The

waiting time distributions of $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ are much smoother without unlocking patients. Both range between 0 to 24 days with 89.25% served in 10 days and 99.41% in 15 days for P_1L_0 and 89% in 10 days and 99.35% in 15 days for P_2L_0.

Table 2 compares the average daily cost, the percentage of time slots that are not used that will be called unused ratio, the average, standard deviation and maximum of waiting times of the three policies. These results as well as the remaining experimental results confirm the results of Section IV: $\mathbf{P_1}$ and $\mathbf{P_2}$ both improve $\mathbf{P_0}$, both $\mathbf{P_1}$ and $\mathbf{P_2}$ have smaller waiting time variation and smaller maximal waiting time than $\mathbf{P_0}$, $\mathbf{P_2}$ has smaller unused ratio but $\mathbf{P_1}$ has shorter waiting time. These results also show that $\mathbf{P_1}$ and $\mathbf{P_2}$ are actually very close with respect to all performance measures as well as waiting time distribution.

Table 2: Performances of $\mathbf{P_0}$, $\mathbf{P_1}$ and $\mathbf{P_2}$ for the base case

	Daily cost	% unused	Waiting time (days)		
			Average	Standard deviation	Maximum
$\mathbf{P_0}$	5.03	8.84	4.67	7.37	35
$\mathbf{P_1L_0}$	4.74	8.84	4.31	3.73	24
$\mathbf{P_2L_0}$	4.77	8.66	4.39	3.77	24

Remark 3: We have discussed the results with the managers of the collaborating hospital. They are very interested in the proposed methodology and intend to change the current way. However, there are several issues they are concerned about: 1) the impacts of this method on the waiting times of other patients; 2) how to implement the strategies if advance CTS cancellation is considered and 3) whether it benefits from using contracts for multiple medical departments. Decisions of real implementation will made based on the results of the above issues. The implementation has also been delayed due to serious financial problems of the hospital, a common situation for most French public hospitals. Cutting cost has become top priority of this hospital while the work of this paper mainly aims at improving service quality to stroke patients. We expect the implementation of the contract-based solutions resumed after the current financial pressure.

C. Sensitivity analysis of unused time slot cost c

This section considers the base case with contract $\boldsymbol{n_0}$ = [1,1,1,1,2,0,0] and performs sensitivity analysis of the six policies under consideration by changing unused time slot cost c in {1, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100}.

Table 3 shows the thresholds $\boldsymbol{L_0}$, $\boldsymbol{L_1}$ and $\boldsymbol{L_2}$ and the service ratio α of the policies. When c

increases, L_0, L_1 and L_2 increase and α decreases leading to less RTS assignments / reservations and hence smaller unused ratio. Local search reduces the thresholds and hence more RTS reservations and more unused time slots.

Table 3: Thresholds and service ratio *v.s.* penalty cost *c*

c	L_0	L_1	L_2	α
1	{5,5,5,5,4,5,5}	{4,6,2,0,0,5,5}	{5,5,5,5,-18,5,5}	0.69
10	{6,6,6,6,5,5,6}	{6,6,3,2,2,1,6}	{6,6,6,0,-1,-2,6}	0.34
15	{6,6,6,6,5,6,6}	{3,4,4,4,2,6,6}	{6,6,2,2,0,6,6}	0.31
20	{7,6,6,7,6,6,7}	{6,6,5,4,3,4,7}	{5,5,5,4,2,3,7}	0.25
30	{7,7,7,7,6,7,7}	{7,7,6,5,5,7,7}	{6,8,5,5,3,7,7}	0.22
40	{8,8,8,8,7,8,8}	{7,7,6,7,6,7,8}	{6,7,7,7,5,8,8}	0.13
50	{9,9,9,9,8,8,9}	{9,10,9,8,6,8,9}	{9,10,9,8,6,8,9}	0.09
60	{9,9,9,9,8,9,9}	{8,8,9,9,7,8,9}	{8,8,9,9,7,8,9}	0.09
70	{10,10,10,10,9,9,10}	{10,11,11,10,8,8,10}	{9,9,10,10,8,9,10}	0.07
80	{10,10,10,11,10,10,10}	{10,9,11,11,9,10,10}	{10,9,11,11,9,10,10}	0.05
90	{11,11,11,11,10,11,11}	{11,11,10,12,10,10,11}	{11,11,10,12,10,10,11}	0.05
100	{12,12,12,12,11,11,12}	{12,12,11,13,11,11,12}	{12,12,11,13,11,11,12}	0.02

Figures 3 to 7 compare the average daily cost, the unused ratio, the average value, the standard deviation and the maximum of patients waiting times of the six policies $\mathbf{P_0}$, $\mathbf{P_1L_0}$, $\mathbf{P_1L_1}$, $\mathbf{P_2L_0}$, $\mathbf{P_2L_2}$ and $\mathbf{P_3}$. The following observations can be made.

For average daily cost, (i) all daily cost increases when *c* increases; (ii) P_3 has the lowest daily cost; (iii) $\mathbf{P_0}$ has the highest daily cost; (iv) $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ have almost the same daily cost; (v) local search improves $\mathbf{P_1}$ and $\mathbf{P_2}$; (vi) daily cost are almost the same for all policies when *c* > 20.

For unused time slots, (i) the unused ratio decreases when *c* increases, (ii) $\mathbf{P_2L_0}$ has the lowest unused ratio, (iii) $\mathbf{P_0}$ and $\mathbf{P_1L_0}$ have the same unused ratio that is slightly higher than $\mathbf{P_2L_0}$, (iv) local search leads to significantly higher unused ratio, (v) $\mathbf{P_3}$ often has the highest unused ratio.

For the average waiting times, (i) the average delay increases from half day to over 6 days when *c* increases, (ii) $\mathbf{P_0}$ has the highest delay, (ii) $\mathbf{P_2L_0}$ has shorter delay than $\mathbf{P_0}$ but slightly longer delay than $\mathbf{P_1L_0}$, (iii) local search significantly reduces the delay, (iv) $\mathbf{P_3}$ often has the shortest waiting time.

For the standard deviation of waiting times, (i) $\mathbf{P_0}$ always has the largest waiting time

variation but it decreases when *c* increases, (ii) the waiting time variations of other policies increase when *c* increases, (iii) $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ have almost the same variation, (iv) local search reduces this variation, (v) $\mathbf{P_3}$ has often the smaller waiting time variation.

For the maximal waiting time, the sensitivity with respect to *c* is almost the same as that of the standard deviation except that the maximal delay of $\mathbf{P_0}$ remains the same.

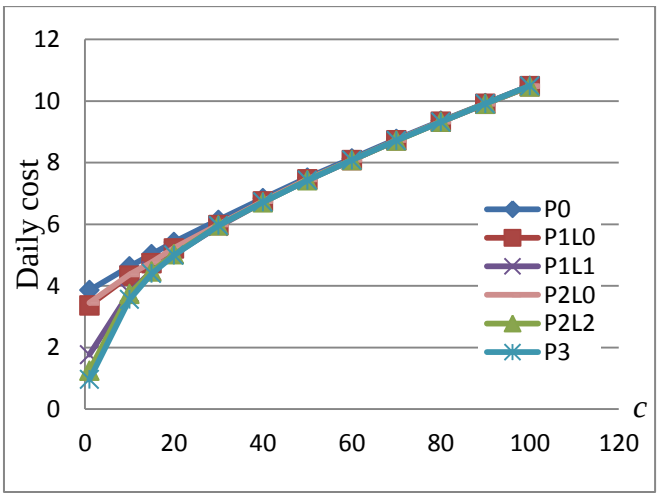

Fig. 3. Daily cost vs. unused time slot cost *c*

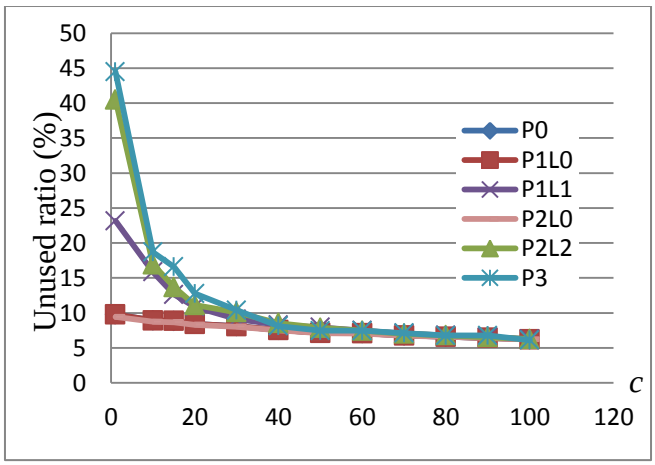

Fig. 4. Unused ratio vs. unused time slot cost *c*

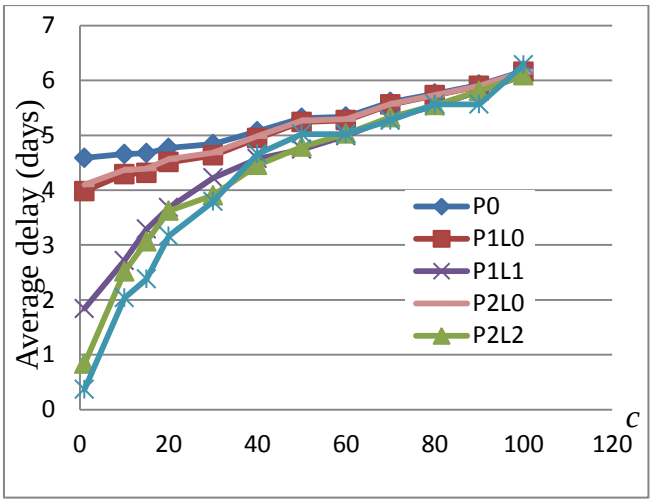

Fig. 5. Average waiting time vs. unused time slot cost *c*

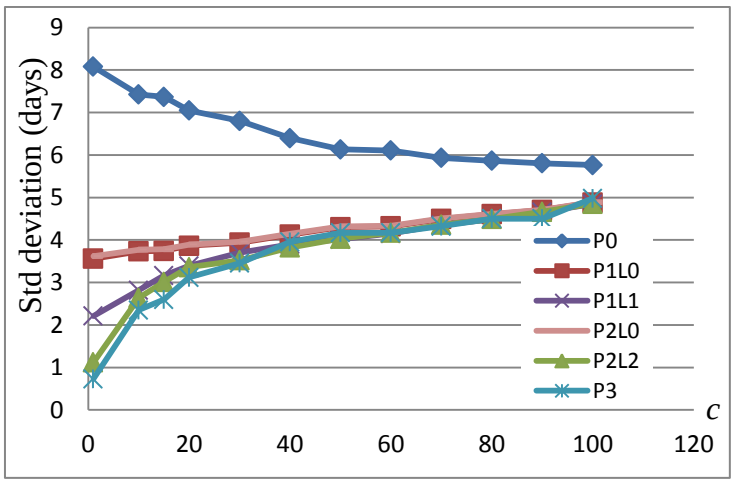

Fig. 6. Waiting time variation vs. unused time slot cost *c*

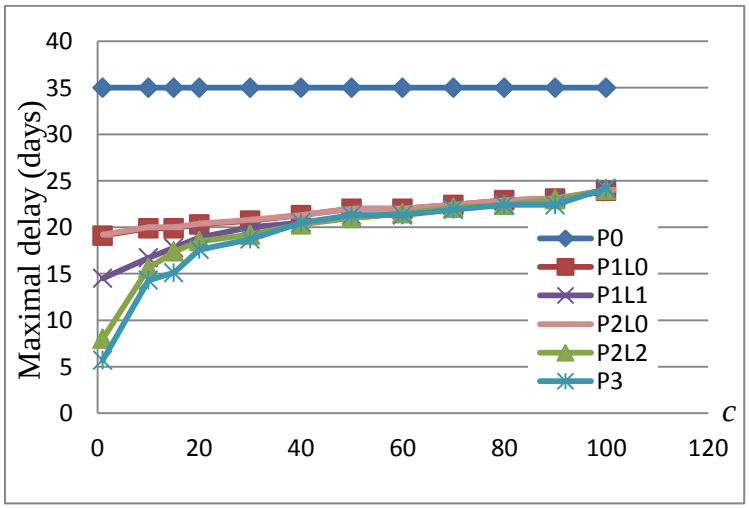

Fig. 7. Maximal waiting time vs. unused time slot cost *c*

D. Sensitivity analysis of RTS delay

This section considers the base case with contract $\boldsymbol{n_0}$ = [1,1,1,1,2,0,0] and performs sensitivity analysis of the six policies under consideration by changing RTS delay T in {2, 10, 20, 30, 35, 40, 50, 60, 70, 80, 90, 100}.

Table 4 shows the thresholds $\boldsymbol{L_0}$, $\boldsymbol{L_1}$ and $\boldsymbol{L_2}$ and the service ratio α of the policies. When T increases, $\boldsymbol{L_0}$, $\boldsymbol{L_1}$ and $\boldsymbol{L_2}$ increase while α is insensitive to the change of T. Local search reduces the thresholds and hence more RTS reservations and more unused time slots.

Table 4: Thresholds and service ratio vs. RTS delays *T*

T	$\mathbf{L_0}$	$\mathbf{L_1}$	$\mathbf{L_2}$	α
2	{3,3,3,3,2,3,3}	{3,3,3,3,2,4,4}	{3,3,3,3,2,3,3}	0.31
10	{4,4,4,4,3,4,4}	{4,3,5,4,3,3,4}	{4,3,5,4,3,3,4}	0.24
20	{5,5,5,5,4,4,5}	{6,3,5,4,3,3,5}	{5,3,5,4,3,2,5}	0.25
30	{6,6,6,6,5,5,6}	{4,2,8,3,2,2,6}	{6,6,7,3,2,1,6}	0.27
35	{6,6,6,6,5,6,6}	{3,4,4,4,2,6,6}	{6,6,2,2,0,6,6}	0.31
40	{7,6,6,7,6,6,7}	{7,7,6,5,4,2,7}	{7,7,6,2,1,-1,7}	0.31
50	{7,7,7,7,6,7,7}	{7,8,4,3,1,7,7}	{7,8,7,-1,-1,7,7}	0.31
60	{8,8,8,8,7,8,8}	{9,9,8,2,1,8,8}	{9,9,8,8,-3,8,8}	0.29
70	{9,9,9,9,8,8,9}	{9,10,10,3,2,1,9}	{9,10,10,9,8,-5,9}	0.29
80	{9,9,9,9,8,9,9}	{10,9,9,3,1,9,9}	{10,9,9,9,-5,9,9}	0.31
90	{10,10,10,10,9,9,10}	{10,10,10,10,2,1,10}	{10,10,10,10,9,-8,10}	0.31
100	{10,10,10,11,10,10,10}	{11,10,1,11,10,11,10}	{10,10,-9,11,10,11,10}	0.31

Figures 8 to 12 compare different performance measures of the six policies. The following observations can be made.

For average daily cost, all daily cost increases when *T* increases. Except for the case *T* = 2, $\mathbf{P_0}$ has the highest daily cost, $\mathbf{P_3}$ has the lowest cost, $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ has almost the same cost. Local search does improve $\mathbf{P_1}$ and $\mathbf{P_2}$ and $\mathbf{P_1L_1}$ and $\mathbf{P_2L_2}$ have almost the same cost. Cost reduction of RTS reservation over RTS assignment become larger when *T* increases.

For unused time slots, the unused ratios of $\mathbf{P_0}$, $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ are almost the same and decreases as *T* increases. $\mathbf{P_3}$ has the highest unused ratio while the unused ratios of $\mathbf{P_1L_1}$ and $\mathbf{P_2L_2}$ are smaller than that of $\mathbf{P_3}$ but significantly higher than that of $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ especially for large *T*. There is no clear trend of unused ratios of $\mathbf{P_3}$, $\mathbf{P_1L_1}$, $\mathbf{P_2L_2}$.

For the average waiting times, (i) the average delay increases when *c* increases, (ii) $\mathbf{P_0}$ has the highest delay, (ii) $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ has almost the same waiting times, (iii) local search significantly reduces the delay especially for large *T*, (iv) $\mathbf{P_3}$ often has the shortest waiting time, (v) $\mathbf{P_1L_1}$, $\mathbf{P_2L_2}$ and $\mathbf{P_3}$ are able to limit the average waiting time below 3.5 days even for very large *T*.

For the standard deviation of waiting times, (i) the waiting time variations increase when *T* increases, (ii) $\mathbf{P_0}$ always has the largest waiting time variation, (iii) the improvement of RTS reservation policies becomes larger for large *T*, (iv) $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ have almost the same variation, (v) local search significantly reduces this variation, (vi) $\mathbf{P_3}$ has the smaller waiting time variation, (vii) $\mathbf{P_1L_1}$, $\mathbf{P_2L_2}$ and $\mathbf{P_3}$ are able to limit the standard deviation below 3.5 days

even for very large *T*.

For the maximal waiting time, the maximal delay increases as *T* increases, $\mathbf{P_0}$ has the worse maximal delay equal to *T*, $\mathbf{P_1L_0}$ and $\mathbf{P_2L_0}$ have almost the same maximal delay that is significantly lower than *T*, $\mathbf{P_1L_1}$, $\mathbf{P_2L_2}$ and $\mathbf{P_3}$ have even smaller maximal delay and are able to limit it below 22 days even for very large *T*.

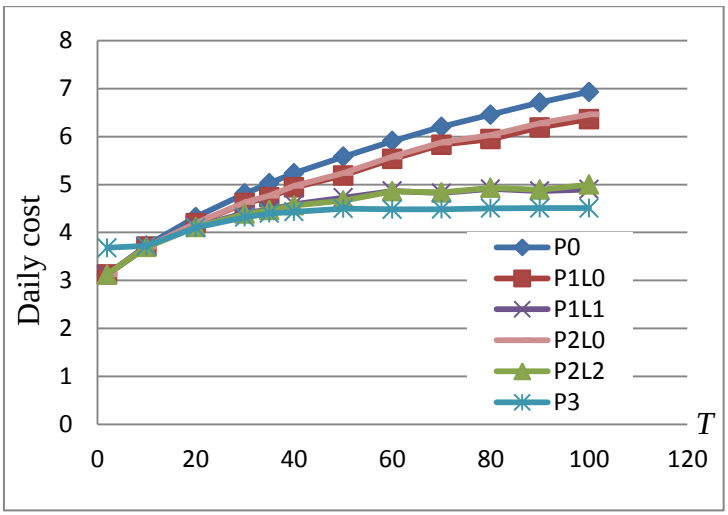

Fig. 8. Daily cost vs. RTS delay *T*

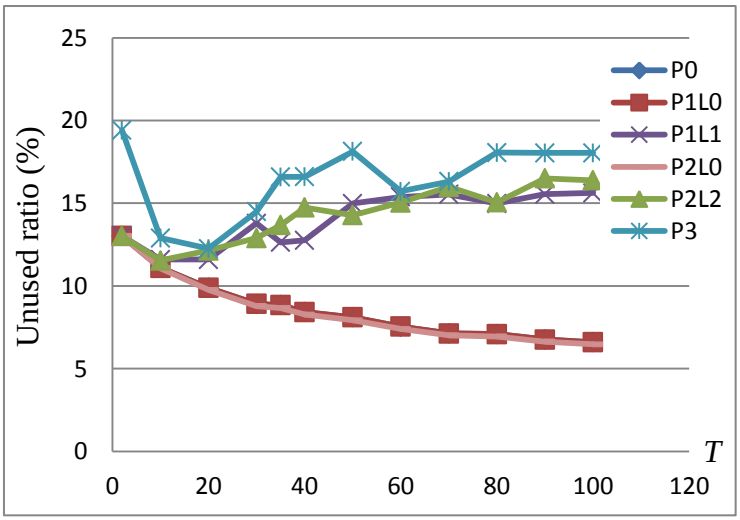

Fig. 9. Unused ratio vs. RTS delay *T*

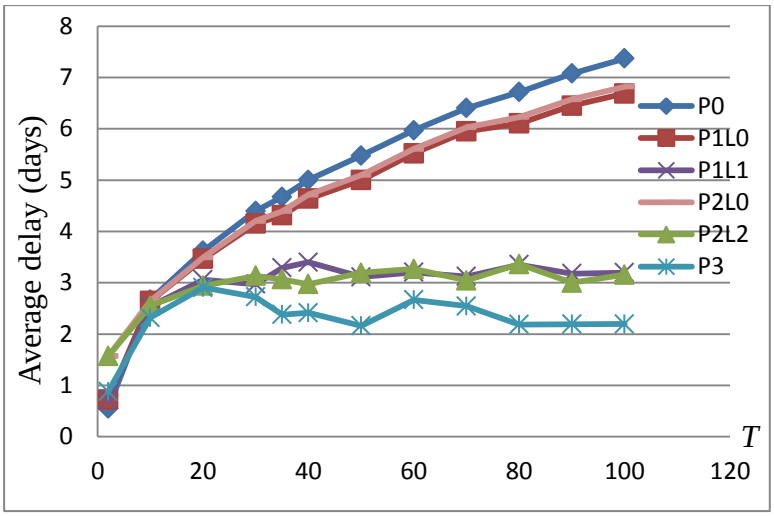

Fig. 10. Average waiting time vs. RTS delay *T*

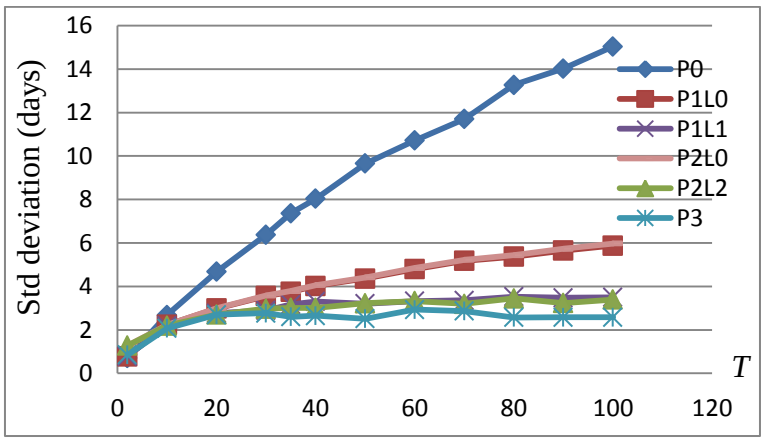

Fig. 11. Waiting time variation vs. RTS delay *T*

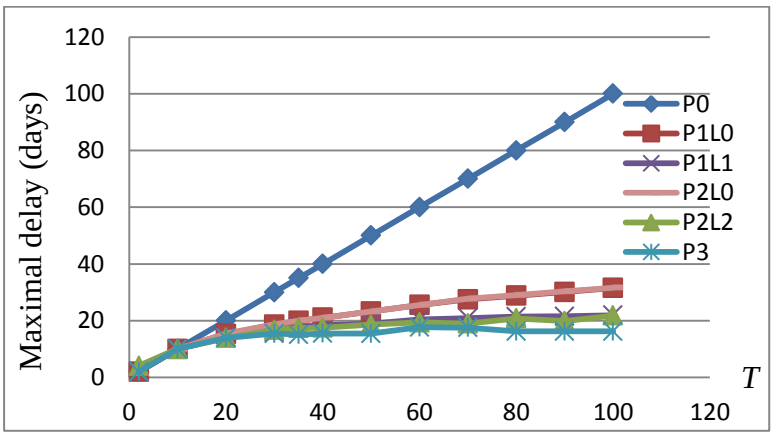

Fig. 12. Maximal waiting time vs. RTS delay *T*

E. Sensitivity analysis of patient arrival rate

This section performs the sensitivity analysis of patient arrival rate by multiplying the patient arrival rates of the base case by 1, 5 and 10. This leads to three cases that are called

respectively Low, Medium, and High demand cases. For all cases, the contract $\boldsymbol{n_0}$ is the optimal contract with RTS assignment determined by the Monte Carlo approach in [1] in order to minimize the average daily cost but without limiting the number of CTS. Local search of $\mathbf{P_1}$ and $\mathbf{P_2}$ is not considered here and the impact is similar to policy $\mathbf{P_3}$.

Table 5 shows the contract $\boldsymbol{n_0}$, the thresholds $\boldsymbol{L_0}$ and the service ratio α of the policies. As the demand increases, $\boldsymbol{n_0}$ and $\boldsymbol{L_0}$ increase while α remains stable.

Table 6 gives the performances of different strategies where each performance for each strategy is given in the order of Low, Medium and High demand cases. The followings are observed. When the arrival rate increases, the daily cost increases but all other performance measures decreases. While the reduction of unused ratio and average waiting time is expected. The reduction of the waiting time variation and maximal waiting time is not straightforward.

RTS reservation strategies do improve the waiting time variations and the maximal waiting time for all cases. This confirms the results of Section III.

Finally, contracts of this section use more CTS than the contract used the previous section which limits the number of CTS to the average weekly arrival rate. With more CTS, the average daily cost, unused ratio and average waiting time are almost the same for all policies. This suggests that the benefits of RTS reservation strategies with respect to these criteria are significant only when the contracted number of CTS is near or below the weekly arrival rate. The optimal contract meets this condition if the unused time slot cost *c* is high.

Table 5: Contracts, thresholds and service ratio vs. arrival rate

Arrival Rate	$\mathbf{n_0}$	$\mathbf{L_0}$	α
Low	{1,1,1,1,3,0,0}	{11,11,11,11,9,10,10}	0.25
Medium	{5,5,5,6,9,1,0}	{21,21,21,21,19,19,21}	0.27
High	{10,9,10,12,17,2,1}	{31,31,31,31,28,29,30}	0.24

Table 6: Performances vs. patient arrival rate

							Waiting time (days)								
	Daily cost			% unused			Average			Standard Dev			Maximum		
P0	4.48	9.78	13.93	18.16	7.71	6.06	2.13	1.14	0.73	2.74	2.12	1.55	35	35	35
P1L0	4.48	9.75	13.88	18.16	7.71	6.06	2.12	1.13	0.72	2.45	1.46	1.06	15.4	9.9	6.9
P2L0	4.48	9.75	13.88	18.16	7.71	6.06	2.13	1.14	0.73	2.45	1.47	1.07	15.4	9.9	6.9
P3	4.47	9.80	13.91	18.32	8.49	6.27	2.09	1.01	0.69	2.40	1.35	1.03	14.9	8.6	6.8

V. CONCLUSION

This paper proposed different strategies to implement a contract-based MRI reservation process for stroke patients. Previous studies proposed methods for determination of the number of CTS and the corresponding RTS assignment control policy. This paper proves that RTS reservation instead of RTS assignment, i.e. reserve RTS time slots instead assign patient directly to RTS reservation, improves the contract-based solution by reducing unused MRI time slots, patient waiting time and waiting time variation. Three RTS reservation strategies are proposed. The first strategy is built on an artificial queue, the second one use actual patient queue information and the third one is service ratio oriented. Apart from the theoretical analysis, an extensive numerical experiment is performed to show the effectiveness of the RTS reservation strategies.

Future research can be pursued in several directions. First, it would be interesting to combine the RTS reservation strategies with advance CTS cancelation. Queueing analysis of RTS reservation strategies is another interesting direction. Other directions are related to the impacts of contracting MRI time slots for the neural vascular department on the waiting times of other patients. Contract design for multiple medical departments and for patients requiring multiple diagnostic facilities is an important challenging issue.

APPENDIX. ANALYSIS OF IMPLEMENTATION STRATEGIES

A. RTS assignment vs. RTS reservation with artificial queue

Property 1: $x_{0,t} \geq x_{1,t}, \forall t \geq 0$.

Property 2: $\bar{U}_1 = \bar{U}_0$.

Property 3: The maximal waiting time under P_1 is smaller or equal to T if the thresholds L_t are such that $L_t \leq \sum_{s=t+1}^{t+T} n_s$ and $L_t - n_{t+1} \leq L_{t+1}$.

Remark 4: Apart from Properties 1-3, the followings holds (i) $y_{1,t} = y_{0,t}$, (ii) $\bar{C}_0 \geq \bar{C}_1$ and (iii) $\bar{W}_0 \geq \bar{W}_1$ where (i) is from the definition of $\mathbf{P_1}$, (ii) from Properties 1 and 2 and (iii) from Property 1 and relation (9). In other words, $\mathbf{P_1}$ not only reduces the maximal waiting time by avoiding "unlucky" patients directed to longer regular reservation but also reduces the average waiting time. □

Remark 5: The conditions of Property 3 basically assume that patients in CTS queue under

$\mathbf{P_0}$ wait no longer than T and a patient in the CTS queue will not be directed to RTS. The last condition holds for any optimal RTS assignment policy. By relaxing the first condition, it is still possible to show that the maximal waiting time of $\mathbf{P_1}$ does not exceed that of $\mathbf{P_0}$. □

Proof of Property 1 by induction: The property clearly holds for $t = 0$. Assume that it holds for day t-1. We now prove that it holds for day t. First, for $\mathbf{P_0}$ and for day t, $q_{0,t-1}$ patients wait for CTS and $x_{0,t-1}$ patients wait for the CTS and RTS. The number of new patients is a_t. During the day, MIN{n_t, $q_{0,t-1}+a_t$} patients are served by CTS and $y_{0,t-T}$ patients by RTS. From the definition of global queue,

$$\begin{aligned} x_{0,t} &= x_{0,t-1} + a_t - \min\left\{n_t, q_{0,t-1} + a_t\right\} - y_{0,t-T} \\ &= \max\left\{x_{0,t-1} + a_t - n_t - y_{0,t-T}, x_{0,t-1} - q_{0,t-1} - y_{0,t-T}\right\} \end{aligned}$$

From definition, we have

$$y_{1,t} = y_{0,t}, \forall t \,. \tag{14}$$

Combining the above with (4) for i=1, (14), the induction assumption and $x_{0,t-1} \geqslant q_{0,t-1} + y_{0,t-T}$, we have $x_{0,t} \geqslant x_{1,t}$ which completes the proof. □

Remark 6: From the proof of Property 1, with the same number of time slots, more patients can be served with RTS reservation $\mathbf{P_1}$ than with RTS assignment $\mathbf{P_0}$. The deviation, i.e. $x_{0,t} > x_{1,t}$, happens when $n_t > q_{0,t-1}+a_t$. In this case, in $\mathbf{P_0}$, there are not enough patients to fill CTS and patients directed to RTS cannot be redirected. In $\mathbf{P_1}$, as no patients are directed RTS, these extra CTS time slots can be filled by patients that were directed to RTS in $\mathbf{P_0}$. □

Proof of Property 2: By definition of global queue,

$$x_{0,t} = x_0 + \sum_{s=1}^{t} a_s - \sum_{s=1}^{t-T} y_{0,s} - \sum_{s=1}^{t}\left(n_s - u_{0,s}\right)$$

$$x_{1,t} = x_0 + \sum_{s=1}^{t} a_s - \sum_{s=1}^{t}\left(n_s + y_{1,s-T} - u_{1,s}\right)$$

As $y_{0,t} = y_{1,t} = 0, \forall t \le 0$, subtracting the two relations leads to:

$$x_{0,t} - x_{1,t} = \sum_{s=1}^{t} u_{0,s} - \sum_{s=1}^{t} u_{1,s}$$

which proves the property if both $x_{0,t}$ and $x_{1,t}$ are finite numbers. The finiteness of $x_{0,t}$ is true as

$q_{0,t} \leqslant L_t$, $y_{0,t} \leqslant q_{0,t-1}+a_t \leqslant L_{t-1}+a_t$, and $x_{0,t} = q_{0,t} + \sum_{s=t-T+1}^{t} y_s \le TL^* + \sum_{s=t-T}^{t} a_s$ where L^* is the maximum L_t for all t. The finiteness of $x_{1,t}$ follows from Property 1. □

Remark 7: From the proofs of Properties 1 and 2, $\sum_{s=1}^{t} u_{0,s} - \sum_{s=1}^{t} u_{1,s} \ge 0$. Policy $\mathbf{P_1}$ has less unused time slots even though the average underutilization cost is the same.

Proof of Property 3: Consider the last patient arriving in day t under $\mathbf{P_1}$, i.e. the $x_{1,t}$-th patient in the patient queue. Assume by contradiction that its waiting time exceeds T. As patients are served in FIFO order,

$$x_{1,t+T} > \sum_{\tau=t+1}^{t+T} a_\tau$$

Consider $\mathbf{P_0}$. From relation (1) and $q_{0,t} \leqslant L_t$, the conditions of the property ensure that all patients in $x_{0,t}$ have been served by time $t + T$. As a result,

$$x_{0,t+T} \le \sum_{\tau=t+1}^{t+T} a_\tau$$

which contradicts Property 1 and concludes the proof. □

B. RTS assignment vs. RTS reservation with real queue

Property 4. $\left(x_{0,t}, y_{0,t}, q_{0,t}\right) \ge \left(x_{2,t}, y_{2,t}, q_{2,t}\right), \forall t \ge 0$.

Property 5. $\left(x_{0,t}, y_{0,t}, q_{0,t}, u_{0,t}\right) = \left(x_{2,t}, y_{2,t}, q_{2,t}, u_{2,t}\right), \forall t < t^*$, $q_{0,t^*} = 0 > q_{2,t^*}$, $x_{0,t^*} > x_{2,t^*}$, $u_{0,t^*} > u_{2,t^*}$ where

$$t^* = \min\left\{t \in \mathrm{IN}^+ : u_{0,t} > 0, Y_0\left(t-T+1, t-1\right) > 0\right\}$$

is the first day such that $\mathbf{P_0}$ has unused CTS and outstanding RTS patients.

Remark 8: From Property 5, $\mathbf{P_0}$ and $\mathbf{P_2}$ deviate when P_0 has both unused CTS and outstanding RTS patients. In such days, $\mathbf{P_2}$ fills these unused time slots with outstanding patients that were directed to RTS in $\mathbf{P_0}$.

Property 6. $D_{2,t} \ge D_{0,t}$, $U_{2,t} \le U_{0,t}$, $\bar{C}_2 \le \bar{C}_0$, $\bar{U}_2 \le \bar{U}_0$, $\bar{W}_2 \le \bar{W}_0$.

Property 7. The maximal waiting time of the reservation process $\mathbf{P_2}$ is smaller or equal to T if

the thresholds L_t are such that $L_t \le \sum_{s=t+1}^{t+T} n_s$ and $L_t - n_{t+1} \le L_{t+1}$.

Remark 9: From Properties 6 and 7, compared with the RTS assignment policy $\mathbf{P_0}$, RTS reservation $\mathbf{P_2}$ with real queue improves both the ratio of unused time slots and the average waiting time of patients. It further avoids unlucky patients with lengthy regular reservation time *T*.

The proofs of the above properties need the following basic relations of policy $\mathbf{P_2}$.

$$u_{2,t} = \left(n_t - q_{2,t-1} - a_t - Y_2\left(t-T+1, t-1\right)\right)^+ \tag{15}$$

$$y_{2,t} = \left(q_{2,t-1} + a_t - n_t - L_t\right)^+ \tag{16}$$

$$q_{2,t} = \max\left(x_{2.t-1} + a_t - n_t - Y_2\left(t-T, t\right), -Y_2\left(t-T+1, t\right)\right) \tag{17}$$

$$q_{2,t} = \max\left(q_{2,t-1} + a_t - n_t - y_{2,t}, -Y_2\left(t-T+1, t\right)\right) \tag{18}$$

where (15) is from (1) and (6), (16) from (4), (12) and (1), (17) from (1) and (4) , (18) from (17) and (1). The proof of property 7 is similar with that of Property 3 and is neglected.

Proof of Property 4 by induction: The property clearly holds for $t = 0$. Assume that it holds up to day t-1. We now prove that it holds for day t.

(A) $y_{0,t} \ge y_{2,t}$ as, by induction and relations (10) and (16),

$$y_{0,t} = \left(q_{0,t-1} + a_t - n_t - L_t\right)^+ \ge \left(q_{2,t-1} + a_t - n_t - L_t\right)^+ = y_{2,t}$$

(B) $q_{0,t} \ge q_{2,t}$ as

$$q_{2.t} \le \left(q_{2,t-1} + a_t - n_t\right)^+$$

$$q_{2,t} \le \left(q_{0,t-1} + a_t - n_t\right)^+$$

$$q_{2.t} \le \min\left(\left(q_{0,t-1} + a_t - n_t\right)^+, L_t\right) = q_{0,t}$$

where the first inequality is from (18) and the fact $y_{2,t} \geqslant 0$, the second from the induction assumption, the third from the second inequality, (13) and (11).

Finally $x_{0,t} \geqslant x_{2,t}$ as

$$x_{2,t} = q_{2,t} + Y_2\left(t-T+1,t\right) \le q_{0,t} + Y_0\left(t-T+1,t\right) = x_{0,t}$$

where the first equality is from (1), the second inequality from (A), (B) and induction, the third equality from (1). □

Proof of Property 5: We first prove $\left(x_{0,t}, y_{0,t}, q_{0,t}, u_{0,t}\right) = \left(x_{2,t}, y_{2,t}, q_{2,t}, u_{2,t}\right), \forall t < t^*$ by induction. It clearly holds for t = 0. Assume that it holds up to day t-1. We now prove that it holds for day t.

(i) Proof of $u_{0,t}$=$u_{2,t}$. By definition of t*, if $u_{0,t} > 0$, then $Y_0(t\text{-}T\text{+}1,t\text{-}1)=0$ and, from (5) and (15), $u_{0,t}$=$u_{2,t}$. Otherwise, from (15), $0 \leqslant u_{2,t} \leqslant (n_t\text{-}q_{2,t\text{-}1}\text{-}a_t)^+ = u_{0,t}$. Since $u_{0,t}$ = 0 in this case, $u_{0,t}$= $u_{2,t}$.

(ii) From (16) and (10), the induction assumption implies $y_{0,t}$= $y_{2,t}$.

(iii) Proof of $q_{0,t}$= $q_{2,t}$. From relations (10), (11) and (18)

$$q_{0,t} = \left(q_{0,t-1} + a_t - n_t\right)^+ - y_{0,t} \tag{19}$$

$$q_{2,t} = \max\left(q_{2,t-1} + a_t - n_t, -Y_2\left(t-T+1,t-1\right)\right) - y_{2,t} \tag{20}$$

If $Y_0\left(t-T+1,t-1\right)=0$, by induction and (ii), $q_{0,t}$=$q_{2,t}$. Otherwise, $u_{0,t}$=0, from (5), $q_{0,t\text{-}1}$+a_t-n_t $\geqslant$0. By induction assumption, (19) and (20), $q_{0,t}$=$q_{2,t}$.

(iv) From (i), induction assumption and (ii) and (iii), $x_{0,t}$=$x_{2,t}$.

Consider now day t*. $u_{0,t^*} = n_{t^*} - q_{0,t^*-1} - a_{t^*} > 0$, $Y_0\left(t^*-T+1,t^*-1\right) > 0$. From (15),

$$\begin{aligned} u_{2,t^*} &= \left(n_{t^*} - q_{2,t^*-1} - a_{t^*} - Y_2\left(t^*-T+1,t^*-1\right)\right)^+ \\ &= \left(n_{t^*} - q_{0,t^*-1} - a_{t^*} - Y_0\left(t^*-T+1,t^*-1\right)\right)^+ \\ &= \left(u_{0,t^*} - Y_0\left(t^*-T+1,t^*-1\right)\right)^+ < u_{0,t^*} \end{aligned}$$

Since $n_{t^*} - q_{0,t^*-1} - a_{t^*} > 0$, from (10), (11) and (1), $q_{0,t^*} = 0$, $y_{0,t^*} = 0$, $x_{0,t^*} = Y_0\left(t^*-T+1,t^*-1\right)$. Similar to the proof of (ii), $y_{2,t^*} = y_{0,t^*} = 0$. From (18),

$$\begin{aligned} q_{2,t^*} &= \max\left(q_{2,t^*-1} + a_{t^*} - n_{t^*}, -Y_2\left(t^*-T+1,t^*-1\right)\right) \\ &= \max\left(q_{0,t^*-1} + a_{t^*} - n_{t^*}, -Y_0\left(t^*-T+1,t^*-1\right)\right) \\ &= \max\left(-u_{0,t^*}, -Y_0\left(t^*-T+1,t^*-1\right)\right) < 0 \end{aligned}$$

Combining with (1), $x_{0,t^*} > x_{2,t^*}$. □

Proof of Property 6: $D_{2,t} \geq D_{0,t}$ is a direct consequence of (7) and $x_{2,t} \leq x_{0,t}$. Since the system is initially empty, $D_{i,t} = N_t + Y_i(1, t-T) - U_{i,t}$ which together with Property 4 implies $U_{2,t} \leq U_{0,t}$ and $\bar{U}_2 \leq \bar{U}_0$. $\bar{W}_2 \leq \bar{W}_0$ flows from (9) and Property 4. $\bar{C}_2 \leq \bar{C}_0$ is a consequence of Property 4, $U_{2,t} \leq U_{0,t}$ and

$$\bar{C}_i = \lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \left(cu_{i,t} + x_{i,t} \right) = \lim_{\tau \to \infty} \frac{1}{\tau} \left(cU_{i,\tau} + \sum_{t=1}^{\tau} x_{i,t} \right) \quad \square$$

C. RTS reservation with artificial queue vs. with real queue

From the definition of $\mathbf{P_1}$ and Property 4,

$$y_{2,t} \leq y_{1,t} \tag{21}$$

By induction and by using (4) and (21), we have

$$x_{2,t} \geq x_{1,t} \tag{22}$$

From Properties 2 and 6, $\bar{U}_2 \leq \bar{U}_1$. From relation (9) and (22), $\bar{W}_2 \geq \bar{W}_1$.

Acknowledgement

The authors acknowledge the support of the National Natural Science Foundation of China (NSFC) under Grant 71131005, 61104173, Specialized Research Fund for the Doctoral Program of Higher Education (20110073120048), and Shanghai Pujiang Program.

REFERENCES

1. Geng, N., Xie, X., Augusto, V., and Jiang, Z., A Monte Carlo Optimization and Dynamic Programming Approach for Managing MRI Examinations of Stroke Patients. IEEE Transactions on Automatic Control, 2011. 56(11): p. 2515-2529.
2. van Sambeek J.R., Joustra P.E., Das S.F., Bakker P.J., Maas M., Reducing MRI access times by tackling the appointment-scheduling strategy. BMJ quality & safety, 2011, 20(12): p. 1075-80
3. Walter, S.D., A comparison of appointment schedules in a hospital radiology department. British Journal of Preventive and Social Medicine, 1973. 27: p. 160-167.
4. Lev, B., Revesz, G., Shea, F., and Caltagirone, R., Patient flow analysis and the

delivery of radiology service. Socio-Economic Planning Sciences, 1976. 10: p. 159-166.

5. Green, L.V., S. Savin, and B. Wang, Managing patient service in a diagnostic medical facility. Operations Research, 2006. 54(1): p. 11-25.
6. Patrick, J., Putmeran, M., Improving resource utilization for diagnostic services through flexible inpatient scheduling: A method for improving resource utilization. Journal of the Operational Research Society, 2007. 58: p. 235-245.
7. Patrick, J., Putmeran, M., Queyranne, M, Dynamic multi-priority patient scheduling for a diagnostic resource. Operations Research, 2008. 56(6): p. 1507-1525.
8. Erdelyi, A., and Topaloglu, H., Computing protection level policies for dynamic capacity allocation problems by using stochastic approximation methods. IIE Transactions, 2009. 41: p. 498-510.
9. Kolisch, R., and Sickinger, S., Providing radiology health care services to stochastic demand of different customer classes. OR Spectrum, 2008. 30(2): p. 375-395.
10. Sickinger, S., and Kolisch, R., A generalized bailey-welch rule and simple tabu search procedure for outpatient appointment scheduling, in Technical Report2008, TUM Business School, Technishce Universitat Munchen.
11. Schutz, H., and Kolisch, R., Capacity allocation for magnetic resonance imaging scanners, Working paper, 2009.
12. Schütz, H.-J. and R. Kolisch, Approximate dynamic programming for capacity allocation in the service industry. European Journal of Operational Research, 2012. 218(1): p. 239-250.
13. Vasanawala, S.S. and T.S. Desser, Accommodation of requests for emergency US and CT: Applications of queueing theory to scheduling of urgent studies. Radiology, 2005. 235(1): p. 244-249.
14. Shumsky, R.A. and F.Q. Zhang, Dynamic Capacity Management with Substitution. Operations Research, 2009. 57(3): p. 671-684.
15. Geng, N., Xie, X.,and Jiang, Z., *Capacity Reservation and Cancellation of Critical Resources.* IEEE Transactions on Automation and Science Engineering, 2011. **8**(3): p. 470-481.
16. Geng, N., Xie, X., *Optimizing contracted resource capacity with two advance cancelation modes.* European Journal of Operational Research, 2012. 221(3): p. 501–512.